Workflow Term Paper BAN440

Contents

[Important dates 2](#_Toc193179537)

[Gathering and cleaning data 2](#_Toc193179538)

[Vinmonopolet 2](#_Toc193179539)

[Population data 2](#_Toc193179540)

[Other data 2](#_Toc193179541)

[Model implementation with R 3](#_Toc193179542)

[Report 4](#_Toc193179543)

# Important dates

|  |  |
| --- | --- |
| **Event** | **Date** |
| Work started | 18.03.2025 |
| Group presentation with feedback | 27.03.2025 |
| Complete project ready for proofing | 07.04.2025 |
| Final submission deadline | 25.04.2025 |

# Gathering and cleaning data

Agder

Akershus

Buskerud – 20 vinmonompolet (180324 willi manuelt)

Finnmark - 10 vinmonompolet (180324 willi manuelt)

Innlandet - 33 vinmonompolet (180324 willi manuelt)

Møre of romsdal - 22 vinmonompolet (180324 willi manuelt)

Nordland – 23 vinmonompolet (180324 willi manuelt)

Oslo – 35 vinmonompolet (180324 willi manuelt)

Rogaland – 24 vinmonompolet (180324 willi manuelt)

Telemark – 13 vinmonompolet (180324 willi manuelt)

Troms - 13 vinmonompolet (180324 willi manuelt)

Trøndelag 34 vinmonompolet (180324 willi manuelt)

Vestfold – 13 vinmonompolet (180324 willi manuelt)

Vestland – 44 vinmonompolet (180324 willi manuelt)

Østfold -11 vinmonompolet (180324 willi manuelt)

As per 18.03 we have a complete data set with information on store names, sales in liters, position (coordinates), municipality number + more.

## Vinmonopolet

Retrieved from:

## Population data

Retrieved from:

## Other data

# Model implementation with R

As per now we have a data set sufficient to estimate a Bresnahan & Reiss model, unless feedback from professors tells us otherwise.

After estimating a model (get a sounder background on the type of model) with only s (population) as the independent variable explaining the number of stores in a municipality. This model is likely not robust enough to use as we get very high thresholds for more than 2 stores. Therefore, we want new variables that possibly can add more depth to the analysis:

We already have

* Sales volume
* Population
* Area
* Density (Population/Area)

We want to add

* Mean income per person
* Grensehandel (people that go on day trips abroad to shop for alcohol)
  + <https://www.ssb.no/varehandel-og-tjenesteyting/varehandel/statistikk/grensehandel>
* Distance from municipality center to closest store
* Age distribution
* Tourism
  + <https://www.ssb.no/statbank/table/12898/>
  + Could only find data on amount of sleepovers with a lot of missing data for some municipalities. It is the best we can do…

# Data Wrangling

### API\_extraction\_stores\_v0

### Dist\_Nearest\_store\_calc

### Kommunesenter\_geodata\_correction

# Report